```python
import pandas as pd
import nltk
nltk.download('punkt')
import wordcloud as w
import matplotlib.pyplot as plt
import sklearn
import warnings
warnings.filterwarnings('ignore')
data="jobs.csv"
df=pd.read_csv(data)
print(df.head())
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
   Unnamed: 0                        Job Salary Job Experience Required  \
0           0    Not Disclosed by Recruiter                    5 - 10 yrs
1           1    Not Disclosed by Recruiter                    2 - 5 yrs
2           2    Not Disclosed by Recruiter                    0 - 1 yrs
3           3      2,00,000 - 4,00,000 PA.                     0 - 5 yrs
4           4    Not Disclosed by Recruiter                    2 - 5 yrs

                                      Key Skills  \
0                   Media Planning| Digital Media
1   pre sales| closing| software knowledge| clien...
2   Computer science| Fabrication| Quality check|...
3                             Technical Support
4   manual testing| test engineering| test cases|...

                               Role Category  \
0                                 Advertising
1                                Retail Sales
2                                         R&D
3   Admin/Maintenance/Security/Datawarehousing
4                         Programming & Design

                               Functional Area  \
0   Marketing , Advertising , MR , PR , Media Plan...
1          Sales , Retail , Business Development
2                     Engineering Design , R&D
3   IT Software - Application Programming , Mainte...
4                     IT Software - QA & Testing

                                 Industry                      Job Title  \
0   Advertising, PR, MR, Event Management  Media Planning Executive/Manager
1         IT-Software, Software Services          Sales Executive/Officer
2               Recruitment, Staffing                     R&D Executive
3         IT-Software, Software Services      Technical Support Engineer
4         IT-Software, Software Services                Testing Engineer

   Unnamed: 8  Unnamed: 9  ...             job Unnamed: 12  \
0         NaN         NaN  ...    Data Scientist         NaN
1         NaN         NaN  ...  Software Engineer         NaN
2         NaN         NaN  ...   Product Manager         NaN
3         NaN         NaN  ...      Data Analyst         NaN
4         NaN         NaN  ...  Software Engineer         NaN

                                   description Unnamed: 14  Unnamed: 15  \
0          Analysing and Interpreting complex datasets         NaN         NaN
1                 Developing  software applications         NaN         NaN
2              Leading product development teams         NaN         NaN
3   Analysing and interpreting data to  inform bus...         NaN         NaN
4                 Developing  software applications         NaN         NaN

   Unnamed: 16  Unnamed: 17  Unnamed: 18  Ratings  Location
0         NaN         NaN         NaN      8.0   chennai
1         NaN         NaN         NaN      3.0   chennai
2         NaN         NaN         NaN      5.0   chennai
3         NaN         NaN         NaN      2.0   chennai
4         NaN         NaN         NaN      7.0   madurai
```

Double-click (or enter) to edit

```python
df.drop_duplicates()
print(df.head())
```

```
   Unnamed: 0                        Job Salary Job Experience Required  \
0           0    Not Disclosed by Recruiter                    5 - 10 yrs
1           1    Not Disclosed by Recruiter                    2 - 5 yrs
2           2    Not Disclosed by Recruiter                    0 - 1 yrs
3           3      2,00,000 - 4,00,000 PA.                     0 - 5 yrs
4           4    Not Disclosed by Recruiter                    2 - 5 yrs

                                      Key Skills  \
```

```
     0                    Media Planning| Digital Media
     1    pre sales| closing| software knowledge| clien...
     2    Computer science| Fabrication| Quality check|...
     3                             Technical Support
     4    manual testing| test engineering| test cases|...

                                Role Category  \
     0                             Advertising
     1                            Retail Sales
     2                                     R&D
     3    Admin/Maintenance/Security/Datawarehousing
     4                    Programming & Design

                                Functional Area  \
     0    Marketing , Advertising , MR , PR , Media Plan...
     1            Sales , Retail , Business Development
     2                       Engineering Design , R&D
     3    IT Software - Application Programming , Mainte...
     4                     IT Software - QA & Testing

                                Industry                         Job Title  \
     0    Advertising, PR, MR, Event Management  Media Planning Executive/Manager
     1            IT-Software, Software Services         Sales Executive/Officer
     2                    Recruitment, Staffing                   R&D Executive
     3        IT-Software, Software Services       Technical Support Engineer
     4        IT-Software, Software Services                 Testing Engineer

        Unnamed: 8  Unnamed: 9  ...               job Unnamed: 12  \
     0        NaN         NaN  ...    Data Scientist         NaN
     1        NaN         NaN  ...  Software Engineer         NaN
     2        NaN         NaN  ...   Product Manager         NaN
     3        NaN         NaN  ...      Data Analyst         NaN
     4        NaN         NaN  ...  Software Engineer         NaN

                                    description Unnamed: 14  Unnamed: 15  \
     0         Analysing and Interpreting complex datasets         NaN          NaN
     1                 Developing  software applications         NaN          NaN
     2                 Leading product development teams         NaN          NaN
     3    Analysing and interpreting data to  inform bus...         NaN          NaN
     4                 Developing  software applications         NaN          NaN

        Unnamed: 16  Unnamed: 17  Unnamed: 18  Ratings  Location
     0        NaN         NaN         NaN      8.0   chennai
     1        NaN         NaN         NaN      3.0   chennai
     2        NaN         NaN         NaN      5.0   chennai
     3        NaN         NaN         NaN      2.0   chennai
     4        NaN         NaN         NaN      7.0   madurai

     [5 rows x 21 columns]
```

```
print(df.isnull().values.any())
df.isna().sum()
jn=df[df["Job Salary"]==' Not Disclosed by Recruiter '].index
df.drop(jn,inplace=True)
print(df.info())
```

```
    True
    <class 'pandas.core.frame.DataFrame'>
    Int64Index: 3223 entries, 3 to 11157
    Data columns (total 8 columns):
     #   Column                 Non-Null Count  Dtype
    ---  ------                 --------------  -----
     0   Unnamed: 0             3223 non-null   int64
     1   Job Salary             3223 non-null   object
     2   Job Experience Required  3222 non-null   object
     3   Key Skills             3222 non-null   object
     4   Role Category          3222 non-null   object
     5   Functional Area        3222 non-null   object
     6   Industry               3222 non-null   object
     7   Job Title              3222 non-null   object
    dtypes: int64(1), object(7)
    memory usage: 226.6+ KB
    None
```

```
print("Top 5 jobs")
df['Job Title'].value_counts()[:5]
```

```
    Top 5 jobs
    Software Developer                          2456
    Associate/Senior Associate -(NonTechnical)   684
    Sales/Business Development Manager           579
    Sales Executive/Officer                      439
    Team Lead/Technical Lead                      332
    Name: Job Title, dtype: int64
```

```python
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
import nltk
nltk.download('punkt')
nltk.download('stopwords')
import pandas as pd

stop_words = set(stopwords.words('english'))

df=pd.read_csv("jobs.csv");

clean_job_titles = []

for title in df['Job Title']:
    tokens = word_tokenize(title)
    filtered_tokens = [word for word in tokens if word.lower() not in stop_words]
    clean_title = ' '.join(filtered_tokens)
    clean_job_titles.append(clean_title)

df['Cleaned Job Title'] = clean_job_titles

print(df)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
       Unnamed: 0                 Job Salary Job Experience Required  \
0               0  Not Disclosed by Recruiter                5 - 10 yrs
1               1  Not Disclosed by Recruiter                 2 - 5 yrs
2               2  Not Disclosed by Recruiter                 0 - 1 yrs
3               3     2,00,000 - 4,00,000 PA.                 0 - 5 yrs
4               4  Not Disclosed by Recruiter                 2 - 5 yrs
...           ...                         ...                       ...
27005       29995  Not Disclosed by Recruiter                9 - 12 yrs
27006       29996  Not Disclosed by Recruiter                 1 - 5 yrs
27007       29997  Not Disclosed by Recruiter                 4 - 6 yrs
27008       29998  Not Disclosed by Recruiter                5 - 10 yrs
27009       29999  Not Disclosed by Recruiter                 2 - 7 yrs

                                             Key Skills  \
0                           Media Planning| Digital Media
1        pre sales| closing| software knowledge| clien...
2        Computer science| Fabrication| Quality check|...
3                                       Technical Support
4        manual testing| test engineering| test cases|...
...                                                   ...
27005    Graphics| C++| USB| Project management| SOC| ...
27006    Service delivery| LMS| CRE| Integration servi...
27007       Counselor| Mentor| Trainer| Advisor| Teaching
27008    Security Analysis| Software Development Life ...
27009    Press release| Public relations| Publishing| ...

                                     Role Category  \
0                                        Advertising
1                                       Retail Sales
2                                                R&D
3        Admin/Maintenance/Security/Datawarehousing
4                              Programming & Design
...                                              ...
27005                          Programming & Design
27006                                    Operations
27007                               University Level
27008                          Programming & Design
27009                       Online/Digital Marketing

                                        Functional Area  \
0           Marketing , Advertising , MR , PR , Media Plan...
1                     Sales , Retail , Business Development
2                              Engineering Design , R&D
3           IT Software - Application Programming , Mainte...
4                               IT Software - QA & Testing
...                                                     ...
27005    IT Software - Embedded , EDA , VLSI , ASIC , C...
27006    ITES , BPO , KPO , LPO , Customer Service , Op...
27007         Teaching , Education , Training , Counselling
27008    IT Software - Application Programming , Mainte...
27009    Marketing , Advertising , MR , PR , Media Plan...

                                        Industry  \
0           Advertising, PR, MR, Event Management
```

```python
df['cleanjob']=clean_job_titles
print(df.head())
```

```
      Unnamed: 0                    Job Salary Job Experience Required  \
0              0   Not Disclosed by Recruiter              5 - 10 yrs
1              1   Not Disclosed by Recruiter               2 - 5 yrs
2              2   Not Disclosed by Recruiter               0 - 1 yrs
3              3      2,00,000 - 4,00,000 PA.               0 - 5 yrs
4              4   Not Disclosed by Recruiter               2 - 5 yrs

                                   Key Skills  \
0                  Media Planning| Digital Media
1     pre sales| closing| software knowledge| clien...
2     Computer science| Fabrication| Quality check|...
3                              Technical Support
4     manual testing| test engineering| test cases|...

                                   Role Category  \
0                                     Advertising
1                                    Retail Sales
2                                             R&D
3     Admin/Maintenance/Security/Datawarehousing
4                            Programming & Design

                                   Functional Area  \
0     Marketing , Advertising , MR , PR , Media Plan...
1            Sales , Retail , Business Development
2                        Engineering Design , R&D
3     IT Software - Application Programming , Mainte...
4                       IT Software - QA & Testing

                                   Industry                        Job Title  \
0     Advertising, PR, MR, Event Management  Media Planning Executive/Manager
1             IT-Software, Software Services         Sales Executive/Officer
2                      Recruitment, Staffing                  R&D Executive
3             IT-Software, Software Services       Technical Support Engineer
4             IT-Software, Software Services                Testing Engineer

                     Cleaned Job Title                          cleanjob
0     Media Planning Executive/Manager  Media Planning Executive/Manager
1              Sales Executive/Officer           Sales Executive/Officer
2                        R & Executive                     R & Executive
3           Technical Support Engineer        Technical Support Engineer
4                     Testing Engineer                  Testing Engineer
```
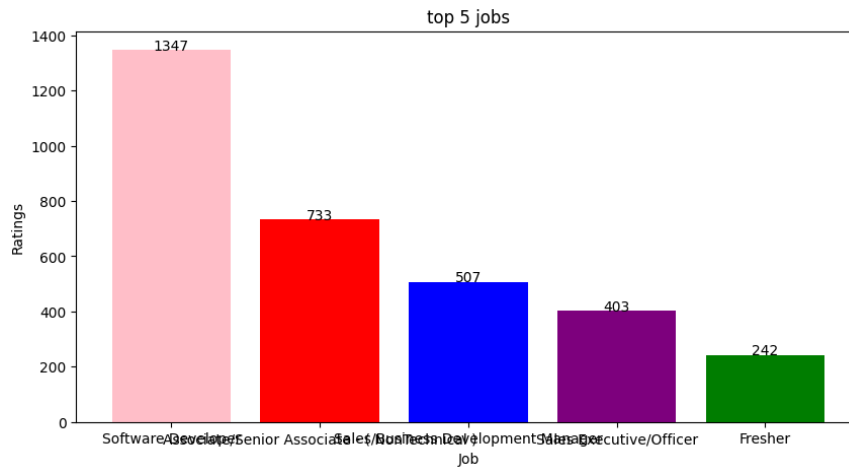
```python
qty=df['cleanjob'].value_counts()[:5].tolist()
label=df['cleanjob'].value_counts()[:5].index.tolist()
print(qty)
print("Top 5 jobs :",str(label))
```

```
    [154, 102, 67, 45, 40]
    Top 5 jobs : ['Software Developer', 'Associate/Senior Associate -(NonTechnical)', 'Sales/Business Development Manager', 'Sales Execu
```

```python
import matplotlib.pyplot as plt
qty=df['cleanjob'].value_counts()[:5].tolist()
label=df['cleanjob'].value_counts()[:5].index.tolist()
def add(x,y):
  for i in range(len(x)):
    plt.text(i,y[i],y[i],ha='center')
sk=label
jq=qty
plt.figure(figsize=(10,5))
plt.bar(sk,jq,color=['pink','red','blue','purple','green'])
add(sk,jq)
plt.title("top 5 jobs")
plt.xlabel("Job")
plt.ylabel("Vacancy")
```

Text(0, 0.5, 'Ratings')


```python
import pandas as pd
df =pd.read_csv("jobs.csv")
key_skills_column = 'Key Skills'
all_skills = []
for skills_str in df[key_skills_column]:
    skills_list = skills_str.split('\n')
    all_skills.extend(skills_list)
skills_counts = pd.Series(all_skills).value_counts()
top_5_skills = skills_counts.head(5)
print("Top 5 skills")
print(top_5_skills)
```

```
    Counselor| Mentor| Trainer| Advisor| Teaching                                                                1
    Help Desk| Customer Service| customer relationship| access controls| Asset Management| Vendor Management     9
    Teaching| Educator| Counselor| Mentor| Trainer| Advisor                                                      9
    marketing executive| distributors| customer profiling| sales strategy| market research| Business Development| sales executive   6
    c++| C| design| MySQL| JavaScript| HTML                                                                      5
    dtype: int64
```

```python
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
data=pd.read_csv("jobs.csv")
text = " ".join(i for i in data["Key Skills"])
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords,
                      background_color="white").generate(text)
plt.figure( figsize=(10,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

```python
text = " ".join(i for i in data["Functional Area"])
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords,
                      background_color="white").generate(text)
plt.figure( figsize=(10,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```python
text = " ".join(i for i in data["Job Title"])
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords,
                      background_color="white").generate(text)
plt.figure( figsize=(10,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```python
from sklearn.feature_extraction import text
from sklearn.metrics.pairwise import cosine_similarity
import pandas as pd
data=pd.read_csv("jobs.csv")
feature = data["Key Skills"].tolist()
tfidf = text.TfidfVectorizer()
tfidf_matrix = tfidf.fit_transform(feature)
similarity = cosine_similarity(tfidf_matrix)
indices = pd.Series(data.index, index=data['Job Title']).drop_duplicates()
```

```
def jobs_recommendation(Title, similarity = similarity):
    index = indices[Title]
    similarity_scores = list(enumerate(similarity[index]))
    similarity_scores = sorted(similarity_scores, key=lambda x: x[::], reverse=True)
    similarity_scores = similarity_scores[0:5]
    newsindices = [i[0] for i in similarity_scores]
    return data[['Job Title', 'Job Experience Required',
                 'Key Skills']].iloc[newsindices]

jt=input("Enter job Title: ")
print(jobs_recommendation(jt))
```

```
    Enter job Title: Software Developer
                                      Job Title Job Experience Required  \
    6249          Sales/Business Development Manager              4 - 5 yrs
    6248                         Software Developer              2 - 5 yrs
    6247  Associate/Senior Associate -(NonTechnical)             5 - 10 yrs
    6246                         Software Developer              1 - 6 yrs
    6245  Associate/Senior Associate -(NonTechnical)             1 - 4 yrs
```